

AD-A073 367

DECISIONS AND DESIGNS INC MCLEAN VA

F/G 5/10

SUBJECTIVE VERSUS STATISTICAL IMPORTANCE WEIGHTS: A CRITERION V--ETC(U)

DEC 78 R S JOHN, W EDWARDS

N00014-76-C-0074

NL

UNCLASSIFIED

| OF |

AD
A073367



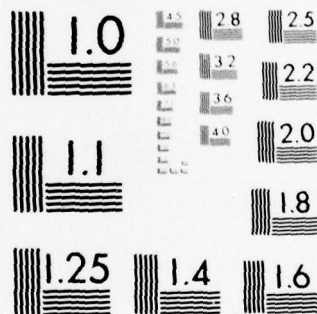
END

DATE

FILMED

10-79

DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A 073367

DDC FILE COPY



USC

UNIVERSITY OF SOUTHERN CALIFORNIA

LEVEL

1

001922-7-T

social science research institute

RESEARCH REPORT

SUBJECTIVE VERSUS STATISTICAL IMPORTANCE WEIGHTS:
A CRITERION VALIDATION

RICHARD S. JOHN
WARD EDWARDS

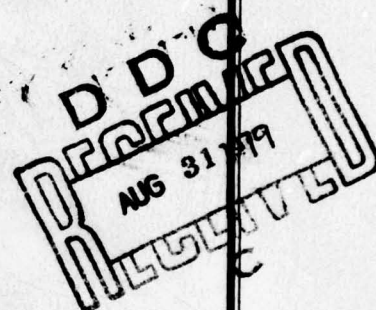
SPONSORED BY:
ADVANCED RESEARCH PROJECTS AGENCY
DEPARTMENT OF DEFENSE

MONITORED BY:
ENGINEERING PSYCHOLOGY PROGRAMS
OFFICE OF NAVAL RESEARCH
CONTRACT No. N00014-76-C-0074, ARPA

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED;
REPRODUCTION IN WHOLE OR IN PART IS
PERMITTED FOR ANY USE OF THE U.S. GOVERNMENT

DECEMBER 1978

SSRI RESEARCH REPORT 78-7



79 08 31 062

28

**Social Science Research Institute
University of Southern California
Los Angeles, California 90007
(213) 741-6955**

INSTITUTE GOALS:

The goals of the Social Science Research Institute are threefold:

- To provide an environment in which scientists may pursue their own interests in some blend of basic and methodological research in the investigation of major social problems.
- To provide an environment in which graduate students may receive training in research theory, design and methodology through active participation with senior researchers in ongoing research projects.
- To disseminate information to relevant public and social agencies in order to provide decision makers with the tools and ideas necessary to the formulation of public social policy.

HISTORY:

The Social Science Research Institute, University of Southern California, was established in 1972, with a staff of six. In fiscal year 1978-79, it had a staff of over 90 full- and part-time researchers and support personnel. SSRI draws upon most University academic Departments and Schools to make up its research staff, e.g. Industrial and Systems Engineering, the Law School, Psychology, Public Administration, Safety and Systems Management, and others. Senior researchers have point appointments and most actively combine research with teaching.

FUNDING:

SSRI Reports directly to the Executive Vice President of USC. It is provided with modest annual basic support for administration, operations, and program development. The major sources of funding support are federal, state, and local funding agencies and private, foundations and organizations. The list of sponsors has recently expanded to include governments outside the United States. Total funding has increased from approximately \$150,000 in 1972 to almost \$3,000,000 in the fiscal year 1978-1979.

RESEARCH INTERESTS:

Each senior SSRI scientist is encouraged to pursue his or her own research interests, subject to availability of funding. These interests are diverse: a recent count identified 27. Four major interests persist among groups of SSRI researchers: crime control and criminal justice, methods of dispute resolution and alternatives to the courts, use of administration records for demographic and other research purposes, and exploitation of applications of decision analysis to public decision making and program evaluation. But many SSRI projects do not fall into these categories. Most project combine the skills of several scientists, often from different disciplines. As SSRI research personnel change, its interests will change also.

①

Subjective versus Statistical Importance Weights:
A Criterion Validation

Research Report 78-7
December, 1978



Richard S. John
Ward Edwards

Social Science Research Institute
University of Southern California

Approved for Public Release
Distribution Unlimited

79 08-31 062

SUMMARY

The present paper proposes a research paradigm for comparing weight estimates to empirically derived "true" weights, thus obtaining a measure of the criterion validity of different weight estimation techniques. Subjects are first taught a multi-attribute utility (MAU) model via multiple-cue probability learning (MCPL) and outcome feedback. Then, various assessments of the importance weight parameters for the model attributes are obtained. Composites formed from these weights are subsequently compared to composites formed from optimal statistical weights derived from outcome feedback.

Data are reported from 17 subjects who were taught one of three "diamond worth" MAU models in 100 feedback trials. The models all involved four attributes (cut, color, clarity, and carat weight), and varied in the "environmental correlations" among the dimensions (either (1) all uncorrelated, (2) one large positive correlation, or (3) two large negative correlations). In addition to the usual MCPL indices of consistency, achievement, and matching, pseudo-matching correlations were computed for weights elicited via the direct subjective procedures of ranking and ratio estimation, the indifference procedures of pricing out and trading off to the most important dimension, and regression weights derived from subjective estimates of the validity coefficients. Overall, the composites formed from the subjects' elicited weights closely corresponded to the "true" weight composites. In addition, a high degree of correspondence was demonstrated among all of the assessed weighting schemes. Individual differences are also reported.

The results of the present study are discussed from both an applied and theoretical perspective. To the decision analyst in the field, the present results give support to the belief that the parameter estimates obtained from clients define a "true" normative preference function. Theoretically, the findings of this study are strong evidence that people are aware of their cognitive processes.

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DDC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or special
A	

TABLE OF CONTENTS

	Page
Summary	i
Acknowledgments	iv
Disclaimer.	v
Tables.	vi
Introduction.	1
Experiment I	
Method	4
Results.	8
Discussion	12
Experiment II	
Method	13
Results.	19
Discussion	25
Conclusions and General Discussion	28
Reference Notes.	34
References	35

ACKNOWLEDGMENTS

This project benefited immeasurably from discussions with Hut Barron, Norman Cliff, Bill Gabrielli, Jim Kahan, Norman Miller, Bob Newman, Dave Seaver, Bill Stillwell, and Detlof von Winterfeldt. Special thanks are given to Jenée Euchner, who provided both technical editing and encouragement. Support for this research came from the Advanced Research Projects Agency of the Department of Defense, contract N00014-76-C-0074, on subcontract P. O. 78-072-0720 from Decisions and Designs, Inc.

DISCLAIMER

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or of the United States Government.

TABLES

Page

Table 1: Sample Attribute Intercorrelations, Validity.	6
Coefficients, and Regression Statistics	
Table 2: Matching (r_m) and Pseudo-Matching Correlations.	11
Table 3: Attribute Intercorrelations, Validity Co-	15
efficients and Regression Statistics	
Table 4: Matching (r_m) and Pseudo-Matching Scores, Ex-	23
periment 2 _m	

Introduction

After several years of research on both subjective weights and statistical weights, considerable controversy over issues of validity exists. Although the literature a decade ago suggested that subjective weights were usually poor (Slovic & Lichtenstein, 1971), recent research has not confirmed this. On the contrary, many studies demonstrating the convergent and criterion validity of subjective weights have appeared (John & Edwards, Note 1). However, influential papers in the field continue to cite the old view that subjects' subjective estimates of attribute importance bear little relationship to reality (e.g., Nisbett & Wilson, 1977).

One of the strongest recent findings is that of Schmitt (1978). He taught his subjects a riskless, additive multi-attribute utility function via outcome feedback in a multiple-cue probability learning (MCPL) setting. Obtaining least-squares regression weights and three different sets of subjective weights, Schmitt compared the composites derived from these weights to those resulting from the "true" regression weights used to generate the outcome feedback. He found that there were absolutely no differences between the matching indices (correlations between composites formed from "true" weights and from subjects' weights) across the four sets of obtained weights. Thus, Schmitt produced hard evidence supporting the accuracy of subjective weights.

Two problems with Schmitt's study deserve mention. First, large positive intercorrelations between attributes were present in all conditions of the experiment. In the face of such serious multi-

collinearity problems, the least-squares regression weights are suspect. While the "true" validity coefficients were all moderately positive (ranging from .42 to .53), the "true" regression weights were non-uniform, and included some negative regression weights (e.g., .63, -.15, .16, .40 for the four-attribute problem). In addition to the problem of determining the "true" weights, the high multi-collinearity presents an even more serious problem. Large positive intercorrelations among dimensions imply that all weighting schemes will yield highly convergent composites. Thus, one is led to suspect that Schmitt (1978) would have had difficulty separating good weights from poor ones, even if he had been able to identify an unambiguous set of "true" weights.

Interestingly, the average subjective weights reported by Schmitt are markedly uniform. The maximum ratio between any pair of weights was about two, and most were essentially equal weighting. It appears that the subjective weights obtained by Schmitt were closer to the validity coefficients than to the least-squares regression weights.

Although our study was designed and performed independently of Schmitt's, the two are natural extensions of one another. We, too, taught subjects a multi-attribute utility function via outcome feedback, and we found that subjects are good at learning weights.

Subjective weights, as well as inferred statistical weights, were compared to the "true" weights derived from the outcome feedback provided. Experiment I is the first comparison of subjective and statistical weights to a "true" model taught under controlled conditions in a context free of interattribute correlations. Experiment II is unique in that the "true" model weights are determined, not through a standard least-squares regression, but by ridge regression. Also,

Experiment II is the first test of an idea, originally proposed by Newman (Note 2), for treating subjective weight estimates as validity coefficients (and not as weight parameters).

Extending Newman's basic idea, Experiment II compared weight parameters based on a ridge regression performed on subjective weight estimates (treated as validity coefficient estimates) with the "true" model weights, derived from a ridge regression on the criterion provided during the outcome feedback trials. In addition, weight elicitation procedures, developed from the axioms of multi-attribute utility theory and not heretofore tested in the MCPL paradigm, were among those employed in Experiment II.

Experiment I

Method

Subjects. Nine undergraduate students at the University of Southern California volunteered for the experiment in partial fulfillment of course requirements in Introductory Psychology. The five males and four females received no other direct compensation or incentive beyond class credits. Subjects were run individually in sessions lasting from 60 to 90 minutes.

Training procedures. Each subject was seated in front of a cathode ray tube (CRT) screen. A standardized cover story was given to each subject, explaining how he/she was to learn, via "computer assisted instruction," the manner in which diamonds are appraised. The subject was told that diamonds could be evaluated on four attributes (cut, color, clarity, and carats) and that each diamond would be presented as a "profile" of ratings (between 0 and 10) on each of the four attributes. The ratings were all related to some physical characteristic of the diamond: cut is determined from a formula for combining certain critical angles and length-to-width ratios obtained from very precise measuring devices; color is determined by examining the diamond under a spectroscope; clarity refers to the number and severity of "inclusions" revealed under a microscope; and carat rating is related to the weight of the stone, such that the smaller the number the lighter the stone. Subjects were informed that in all cases higher attribute ratings were better than lower ones.

After an explanation of how the training would proceed and instruction on how to operate the response keyboard connected to the CRT screen,

subjects began the training phase of the experiment. The entire training phase was controlled by a computer program. Subjects first saw a "diamond profile," presented in the following format:

CUT	COLOR	CLARITY	CARAT
8.6	5.4	8.9	2.1

The program then prints the prompt (PRICE?), and waits for the subject to estimate the price of the diamond. After a number has been properly entered (via the keyboard), the program informs the subject of the "true" price of the diamond (outcome feedback) and how much over or under the estimate is. The program stores the subject's response, clears the screen, and presents the next diamond profile. In all, each subject saw 100 such diamond profiles and outcome feedback.

MAU model. The attribute values specified on the 100 diamond profiles were generated independently from a uniform density function with endpoints 0 and 10. Thus, the expected value of the mean rating on each of the attributes is 5, and the expected variance is about 8.3; also, the expected value of the intercorrelation among the attributes is 0. Since the same "seed" was used to start the random-number generator subroutine for each subject, all subjects saw the same 100 diamond profiles and received the same outcome feedback. Sample means, variances, attribute intercorrelations, validity coefficients, and least-squares regression weights, based on the profiles and feedback provided during the 100 learning trials, are presented in Table 1. As is evident, the sample means, variances, and intercorrelations of the four attributes are very nearly the same as their expected values.

Table 1
Sample Attribute Intercorrelations, Validity Coefficients,
and Regression Statistics
Experiment I

	Attribute Intercorrelation				Validity Coefficient	Regression Statistics	
	CUT	COLOR	CLARITY	CARAT	Price	OLS beta ($\hat{\beta}$)	$r \cdot \hat{\beta}$
CUT	7.7	-.16	-.08	-.07	.36	.46	.17
COLOR		9.1	.01	-.01	.03	.11	.00
CLARITY			7.4	-.13	.09	.24	.02
CARAT				7.7	.84	.90	.75
Price					149×10^4		
	4.9	5.1	5.0	5.1	4259		

Note. Variances are given along the diagonal of the intercorrelation matrix, and means are listed across the bottom.

The outcome feedback used to train the subjects was generated from the following model:

$$\begin{aligned} \text{TRUE PRICE} = & 200 \cdot \text{CUT} + 50 \cdot \text{COLOR} + 100 \cdot \text{CLARITY} + \\ & 400 \cdot \text{CARAT} + 500 + 300 \cdot N(0,1) \end{aligned} \quad (1)$$

where $N(0,1)$ is normal random error with mean 0 and variance 1. The expected value of the mean price is 4250, and the expected total variance is about 186×10^4 . Since the expected error variance is only 9×10^4 , the expected value of the multiple correlation, R_e , is .98 ($\sqrt{(186-9) / 186}$). The sample values of the price mean and variance, given in Table 1, are all quite close to their expected values, as is the sample value of R_e ($= \sum r \cdot \hat{\beta} = .97$). Since the attribute variances are approximately equal, and the attribute intercorrelations are close to zero, the ordinary least-squares (OLS) betas given in Table 1 are roughly proportional to the attribute weights defined in Equation 1.

Direct subjective weight assessment. After completing 100 learning trials, the subject was led into an adjoining room and subjective weights were assessed. Two procedures were used. First, the subject was simply asked to rank-order the attributes from most important to least important in determining overall diamond worth. Next, ratio weights were elicited using Edward's SMART procedure. The least important attribute (identified from the rank-ordering) was assigned a weight of 10, and weights on the other three attributes were determined by the subject. The subject was instructed to make sure that the ratio of any pair of importance weights reflected the number of times more important one attribute was than the other. The ratio weights were simply normalized to sum to one.

Results

Achievement. The correlation between a subject's responses and "true" diamond prices (provided in outcome feedback) is called "achievement" (r_a). It is useful to examine the achievement scores as an indication of the extent to which subjects' knowledge of the model, gained through outcome feedback, was reflected in his/her holistic evaluations. Every subject improved substantially from the first block of 50 trials to the second block of 50 trials. The median value of r_a increased from .68 to .76. It should be noted that had a subject simply responded with numbers proportional to the sum of the four attribute values (equal weighting), a score of .73 would have resulted for r_a . Also, had a subject simply responded with numbers proportional to the value of the diamond on the most important attribute (CARAT), and ignored the other three attributes, he/she would have received a score of .84 for r_a .

Consistency and pseudo-consistency. For each block of 50 trials, a standard multiple regression was performed on each subject's holistic evaluations, using the four attributes as the "predictor" variables. The regression weights derived represent estimates of subjects' importance weight parameters, as was discussed earlier. For each of the 100 stimulus diamonds, composite estimates of worth were formed by applying the subjects' regression weights, ratio weights, and rank weights. The consistency index (r_s) is the adjusted correlation between the composites formed from the subjects' regression weight model and the holistic evaluations of the subject. (Wherry's shrinkage formula was applied to

the obtained \underline{R}^2 to correct for the usual inflated multiple correlations.) Pseudo-consistency is the correlation between direct subjective weight models (ratio and rank) and the holistic evaluations of the subject.

There were three important results regarding consistency and pseudo-consistency. First, the models derived from the three weight estimates were all more consistent with holistic choices over the last half of the training session than over the first half. This increase was especially true for the regression-weight model, where the median adjusted \underline{r}_s changed from .68 to .80. The effect was smaller for the two subjective weight models: the median for the ratio-weight model increased from .65 to .70, and the median for the rank-weight model increased from .68 to .72.

Second, the consistency scores over the last block of trials (median = .80) were substantially larger than the pseudo-consistency scores over the last block (ratio median = .70, rank median = .72). This result is in part explained by the uniformly low pseudo-consistency scores over the last 50 trials by Subjects #7, 8, and 9. Neither the ratio nor rank weights elicited from these three subjects were consistent with the weighting policy used in making the holistic evaluations during the last 50 trials.

The third main result was the near equivalence of pseudo-consistency scores obtained with the ratio-weight model and with the rank-weight model. Apparently, the subjects' weighting policy used in making holistic evaluations is as well described by their subjective rankings of the attributes as by their ratio estimates of attribute importance.

Criterion validity: Matching and pseudo-matching. To assess the criterion validity of each of the three sets of weights, composites formed from each (the same as those discussed in the previous section under convergence) were correlated with composites formed from the "true model" weights, determined from an OLS regression analysis of the outcome feedback (given in Table 1). These correlations, presented in Table 2, are the usual "matching" indices used in MCPL research. The term "pseudo-matching" has been used to describe the correlations involving composites formed from direct subjective weight assessment (ratio and rank), since "matching" is traditionally reserved for the model derived from a regression analysis of holistic choices.

In general, all of the matching and pseudo-matching indices were quite high: models derived from subjects' judgments, whether holistic evaluations (regression weights) or direct assessments (ratio and rank weights), were in good agreement with the "true" multi-attribute utility model. Virtually all of the scores are greater than those obtained with either of the simple heuristic models ("Equal" weights median = .76, "Extreme" weights median = .86). There is some indication that the subjects' statistical regression-weight model (median = .97) is better than the ratio-and rank-weight models (both medians = .94), but these differences appear slight. Most of these differences can be attributed to the inferiority of the direct assessments of Subjects # 7, 8, and 9. As was discussed earlier, the ratio and rank weights elicited from these three subjects were not consistent with their holistic choices.

Table 2
Matching (r_m) and Pseudo-Matching Correlations
Experiment I

Weighting Scheme	Subject									Median	Minimum	Maximum
	1	2	3	4	5	6	7	8	9			
Regression (r_m)	1.00	.94	.98	.98	.93	.97	.97	.97	.94	.97	.93	1.00
Ratio	.97	.94	.91	.99	.97	.96	.90	.85	.71	.94	.71	.99
Rank	.97	.92	.97	.97	.94	.94	.87	.85	.69	.94	.69	.97
Equal										.76		
Extreme (Carat)										.86		

Note: The intercorrelations were computed over all 100 learning trials. Each score represents the correlations between the composites from the weighting scheme and the "true" model composites.

Discussion

Experiment I was designed to test the validity of three procedures for assessing importance weights in the most simple multiattribute situation imaginable: four uncorrelated attributes that combine to determine virtually all of the variance in the hypothetical "overall utility" of the object. The MCPL paradigm provides a standard, or "true" multiattribute utility function against which assessed weight parameters were compared. In general, results indicate that all three weighting schemes are consistent with holistic evaluations, convergent with one another, and closely match the "true" MAU model taught via outcome feedback.

An idiographic analysis suggests that individual differences are present, and that the two direct methods for obtaining importance weights (ratio and rank), were not valid for three of the nine subjects. The relatively high level of achievement obtained by these three subjects, as well as their high consistency and matching scores for the regression weights model, suggest that they did learn the MAU diamond model given in Equation 1. Apparently, these three subjects were either unaware of their learned subjective model for diamond worth, or did not understand the instructions for ratio- and rank-weight assessment. Neither of these alternative explanations is palatable, however. The achievement, consistency, and matching indices were simply too high to justify unawareness, and there is not much to misinterpret in the instructions to "rank-order the attributes from most important to least important."

It is intriguing to note, post hoc, that the three subjects in question are all female; thus, although all five male subjects gave valid subjective importance weights, only one out of four females did so.

Interestingly, all three ranked CUT as the most important attribute. Since CUT was also the second most important attribute in the "true" MAU model, this agreement is ambiguous. Subjects #7, 8, 9 could have been expressing a common fact about real diamonds, or they could have simply "come close" in their direct estimate of the most important attribute. Obviously, more data are required before these sex-difference speculations can be resolved.

Experiment II

Method

Subjects. Eight undergraduate students at the University of Southern California volunteered for the experiment. The seven males and one female (chosen without knowledge of Experiment I results) received class credits to fulfill requirements in Introductory Psychology and received no other compensation. Subjects were run individually in sessions lasting from 60 to 90 minutes.

Training procedures. All procedures during the training phase of Experiment II (with the exception of the composition of the programmed MAU model) were identical to those in the first experiment.

MAU models. Two different additive MAU models, each utilizing the four "C" attributes from the first experiment, were used to generate the diamond profiles and corresponding outcome feedback. Half of the subjects saw profiles and feedback from Model "P", which involved a rather large positive correlation between COLOR and CLARITY; the other half were trained on Model "N", which presented diamond profiles with rather large negative correlations between COLOR and CLARITY and between CUT and CARAT.

For model P, three of the four attributes (all but CLARITY) were generated independently from a uniform density function with endpoints

0 and 10. Values on the CLARITY attribute were generated as a function of COLOR and normally distributed random error ($CLARITY = COLOR + 2 \cdot N(0,1)$). Instances in which the value of CLARITY would have been negative or greater than ten were discarded, and an entire new profile was generated. Thus, the expected value of the mean rating on all four attributes is 5. Had no profiles been discarded, the expected value of the variances would have been 8.3 for all attributes except CLARITY, which would have been about 12.3. Since some were discarded, the true expected variances are unknown. The expected value of the attribute intercorrelations is zero, except, of course, for that between COLOR and CLARITY. Although one might expect the correlation to be high, the exact value is unknown, since the calculation involves the expected value of the attribute variances. As is evident in the top portion of Table 3, all of the attribute sample means and intercorrelations, based on the 100 profiles, are close to their expected values. The sample intercorrelation between COLOR and CLARITY is .86, and all of the attribute variances are reasonable.

The outcome feedback used to train the four Model P subjects was generated from the following model:

$$TRUE\ PRICE = 0 \cdot CUT + 60 \cdot COLOR + 20 \cdot CLARITY + 40 \cdot CARAT + 100 + 200 \cdot N(0,1) \quad (8)$$

The expected value of the mean price is 700. Since the formulae for the expected price variance and the multiple correlation both require the expected values of the attribute variances, their values are unknown. The sample values of the price mean and variances, along with sample values of validity coefficients and OLS regression weights, are given

Table 3

Attribute Intercorrelations, Validity Coefficients and Regression Statistics

Experiment II

Attribute Intercorrelation				Validity Coefficient		Regression Statistic	
CUT	COLOR	CLARITY	CARAT	Price	OLS beta ($\hat{\beta}$)	Ridge Weight	$r \cdot \hat{\beta}$
Model P							
CUT	9.1	-.01	-.02	.13	.08	.07	.01
COLOR		8.5	.86	.76	.79	.52	.60
CLARITY			10.3	.60	-.02	.17	-.01
CARAT				.26	.28	.26	.07
Price				96,613			
	4.8 ^a	5.1	5.1	4.8			
Model N							
CUT	5.8	.06	-.04	-.23	.07	.05	-.02
COLOR		7.8	-.95	.73	.79	.63	.58
CLARITY			6.6	-.69	.02	-.13	-.01
CARAT				.32	.46	.43	.15
Price				66,362			
	4.7 ^a	5.0	4.9	5.3			

Note: Variances are given along the diagonals of each intercorrelation matrix.

^aMeans are given in this row.

at the top of Table 3. The square root of the sum of the products of \underline{r} and $\hat{\beta}$, .82, is the sample value of the environmental multiple correlation, R_e .

Because of the high multi-collinearity between COLOR and CLARITY, one might suspect that the inverse of the attribute (predictor) matrix is ill-conditioned. The observation of an eigenvalue of .13 provided confirmation. With small eigenvalues in the inverse of the predictor matrix, major discrepancies between the OLS regression weights and the "true" population weights are virtually guaranteed. Ridge regression was applied to the sample attribute intercorrelations and validity coefficients displayed in the top portion of Table 3. A "ridge trace" was generated, and the constant value (.2) added to the diagonal of the correlation matrix was chosen at that point in the trace where the betas seemed to stabilize. The ridge regression weights are also presented at the top of Table 3. They yield a sample multiple correlation of .81, only slightly less than that for OLS weights. As can be seen, these weights are strikingly different from the OLS regression weights. In particular, the sign of the CLARITY weight, negative for the OLS analysis, is positive for the ridge analysis. Also, the magnitude of the COLOR weight has decreased substantially (from .79 to .52). In general, the ridge regression weights are much closer to the validity coefficients than are the OLS weights.

An analogous procedure was followed for Model N. Here, two of the attributes (CUT and COLOR) were generated independently from a uniform density function with endpoints 0 and 10. Values on the CLARITY attribute were generated as a function of COLOR and normally distributed random error ($CLARITY = 10 - COLOR + N(0,1)$); CARAT was

generated from CUT and normally distributed random error ($CARAT = 10 - CUT + 2 \cdot N(0,1)$). As for Model P, any profile with a value on CLARITY or CARAT outside the 0 to 10 range was discarded and a new profile was generated. Thus, the expected value of the mean rating on all four attributes is 5, and the expected values of the variances are again unknown, due to the discarding of some generated profiles. Had no profiles been discarded, the expected variances would have been about 8.3 for CUT and COLOR, about 9.3 for CLARITY, and about 12.3 for CARAT. The expected value of the attribute intercorrelations is zero, except for that between CUT and CARAT and between COLOR and CLARITY. Although these two correlations are expected to be negative, calculation of their exact value requires the expected values of the attribute variances, which are unknown. The sample attribute means and intercorrelations, presented in the bottom portion of Table 3, are all close to their expected values. The sample intercorrelation between COLOR and CLARITY is $-.95$, and that between CUT and CARAT is $-.74$. Overall, the sample attribute variances are lower than those for Model P.

The outcome feedback was generated from the following model:

$$TRUE\ PRICE = 30 \cdot CUT + 80 \cdot COLOR + 10 \cdot CLARITY + 60 \cdot CARAT + 300 + 150 \cdot N(0,1) \quad (9)$$

The expected value of the mean price is 1200, and the expected variance and expected multiple correlation are both unknown, since they depend upon the unknown expected attribute variances. The sample price mean and variance, the validity coefficients, and the OLS regression weights are given at the bottom of Table 3. The model multiple correlation, R_e , is $.84$ ($= \sqrt{\Sigma r \cdot B}$).

Inspection of the eigenvalues of the inverse of the attribute (predictor) matrix yielded strong evidence for ill-conditioning and OLS mis-estimation, the smallest eigenvalue being less than .05. A ridge regression analysis was applied to the attribute intercorrelations and validity coefficients displayed at the bottom of Table 3. Again, the critical constant (.2) added to the diagonal of the intercorrelation matrix of attributes was determined from an inspection of the "ridge trace." The ridge weights, given at the bottom of Table 3, yield a multiple correlation of .83 (very close to the .84 value for the OLS weights). Although the ridge weights are ordinarily equivalent to the OLS weights, they are different in sign on two attributes. The ridge analysis suggests that CLARITY should have a negative orientation to overall Price, consistent with the validity coefficient. In general, the ridge weights are closer to the validity coefficients than are the OLS regression weights, just as was the case for Model P.

Direct subjective weight assessment. As in the first experiment, subjects were led into an adjoining room, and rank and ratio weights were assessed. Two additional procedures were employed after ratio-weight assessment: "pricing out" and "trading off to the most important dimension" (Keeney and Raiffa, 1976). For the trade-off procedure, subjects essentially specify the change on the most important dimension that is equivalent to a standard change on each of the other three dimensions. For the pricing-out method, subjects must specify an amount of money that is equivalent to a standard change on each of the four attributes. For all four assessment techniques, subjects were forced to be consistent about their implied attribute rankings. The reasoning behind all inconsistencies was explained. Notwithstanding possible Rosenthal effects, all subjects expressed a desire to change

their responses to alleviate the problem. Only three instances of inconsistency, all minor (weights were very close in magnitude), were observed.

Results

Achievement. Achievement scores (correlations between subjects' holistic responses and outcome feedback) for each of the two blocks of fifty trials were calculated. As in the first experiment, r_a showed a consistent increase (median increased from .62 to .73) for all four Model P subjects (P-1, P-2, P-3, P-4). However, the four Model N subjects (N-1, N-2, N-3, N-4) showed no stable pattern for r_a scores. Although two of the subjects' r_a scores remained about the same (N-2 and N-4), Subject N-1 showed a drastic decrease (from .39 to .06) while Subject N-3 increased substantially (from .38 to .69).

Just as in Experiment I, Model P subjects' unaided holistic evaluation were no better than two simple heuristic strategies, equal weighting and extreme weighting. For Model P, the equal-weighting model correlated .74 with the outcome feedback, and the extreme-weighting model (attending only to COLOR) correlated .76. The Model N subjects, however, performed substantially worse than the extreme-weighting model (COLOR only). Although the equal-weighting model only correlated .28 with the Model N feedback, one subject performed even worse during the fifty-trial block. Given the somewhat lower predictability of the Model P feedback as compared to that for Model N (see Table 3), and their near equivalence in predictability for the simple extreme-weighting heuristic, the clear differences in achievement between Model P and Model N subjects are surprising.

Consistency and pseudo-consistency. As in the first experiment, consistency studies (adjusted multiple correlation from OLS regression analysis on subjects' holistic responses) were obtained over each trial block. In addition, consistency scores were computed using ridge weights (constant added to diagonal was .2 for Model P and .3 for Model N) instead of the OLS weights. Pseudo-consistency scores were computed for ratio, rank, price-out, and trade-off weights over both trial blocks. Both "OLS-ratio" and "ridge-ratio" weights were obtained via regression analysis using the elicited ratio weights as estimates of the validity coefficients, and pseudo-consistency scores were computed for these two weighting schemes over both trial blocks. (For ridge-ratio weights, the constant added to the diagonal was .2 for Model P and .4 for Model N.) The obtained consistency and pseudo-consistency scores are measures of the degree to which the various weighting schemes yielded composites consistent with subjects' holistic choices.

Several important results are evident here. First, the various MAU models elicited from Model P subjects are much more consistent with their holistic responses from the last block of trials than from the first block. For Model P, the median correlations range from .78 to .88 for the second block, compared to the .73 to .81 median range for the first block. A different pattern emerged for Model N subjects, who showed no differences in consistency from the first trial block to the second.

A second main result is the rather substantial difference between the two models (N and P) in the overall levels of consistency. The maximum median consistency (or pseudo-consistency) score reported for Model N, over both trial blocks and all eight sets of obtained weight

estimates is .63. In contrast, Model P median consistency (or pseudo-consistency) scores are in the 70's for the first trial block and in the 80's for the second. The lower consistency scores for Model N subjects indicates that their holistic responses were less predictable from the four attributes than were those for Model P subjects. The lower pseudo-consistency scores for Model N indicates that Model P subjects were better able to describe the weighting policy they actually used in generating their holistic estimates.

Perhaps the most striking result is the near equivalence among all of the consistency and pseudo-consistency indices for each particular trial block and model. Other than the marked failure of the hybrid OLS-ratio technique for Model N, all of the assessed (or derived) weighting schemes predicted subjects' holistic responses equally well; little or no consistent pattern emerged from the data. Although the OLS and ridge composites tended to be in closer correspondence to holistic responses than composites from either the direct subjective assessments (ratio and rank) or the indifference assessments (price-out and trade-off), the differences appear very slight. While ridge and OLS weights are essentially equivalent in power to predict holistic responses, the ridge-ratio weights are substantially more predictive than the OLS-ratio weights. Thus, when the attribute validity coefficients were derived from subjects' holistic evaluations, little difference between the ridge and OLS weights emerged. However, when the validity coefficients were estimated directly (from the subjects' ratio-weight assessments), the ridge analysis yielded weights strikingly more predictive of holistic responses.

Criterion validity - matching and pseudo-matching. The criterion validity of each of the eight sets of subjects' weights was assessed by computing matching indices for OLS and ridge weights and pseudo-matching indices for the remaining weighting schemes. Both the "true OLS" and "true ridge" weights, presented in Table 3 for models P and N, were used as criterion models. The matching and pseudo-matching correlations are presented in Table 4.

Overall, the matching and pseudo-matching measures are quite high for Model P subjects. Median correlations range from .94 to 1.00 for Model P subjects, indicating that the assessed (derived) weight composites agree with both sets of true weight composites. For Model P, there is no evidence of differences among the eight sets of assessed weights or between the two sets of "true weights." Although the two hybrid weight composites diverge somewhat from the "true weight" composites for subject P-1, all four Model P subjects display the same general pattern of extremely high matching and pseudo-matching.

The pattern of results is more complicated for the four Model N subjects. Although the correspondence is lower, in general, than that for Model P, there are obvious individual differences. In comparing the subjects' ridge weights to their OLS weights, all four subjects show better pseudo-matching for their ridge weights when the validity coefficients are directly estimated from the ratio weight assessment. Two of the subjects (N-1 and N-2) show a superiority for ridge weights when the validity coefficients are estimated from the subjects' holistic evaluations. As for Model P, there is no consistent pattern in the pseudo-matching scores for the four post-training sets of weights. For

Table 4
Matching (r_m) and Pseudo-Matching Scores
Experiment II

Weighting Scheme	Subject							Median	
	P-1	P-2	P-3	P-4	N-1	N-2	N-3	N-4	Model P Model N
Ridge regression	.94 ^a	.99	.99	.98	.47	.81	.99	.92	.99 .87
	.91 ^b	.98	.97	.96	.51	.84	.97	.92	.97 .88
OLS regression	.94	.99	.99	.97	.28	.69	.99+	.87	.98 .78
	.91	.99	.97	.95	.37	.73	.98	.91	.96 .82
Ridge-ratio	.86	.99+	.99	.99	.44	.60	.74	.42	.99 .52
	.83	.99+	.99+	.98	.55	.70	.82	.53	.99 .63
OLS-ratio	.81	.99	.90	.98	.40	.38	.44	.30	.94 .39
	.77	.99+	.95	.99	.52	.50	.55	.41	.97 .51
Ratio	.97	.98	.97	.97	.67	.91	.97	.60	.98 .79
	.95	.97	.94	.94	.75	.95	.99	.67	.95 .85
Rank	.90	.99	.99	.99	.81	.91	.91	.45	.99 .86
	.89	.98	.96	.96	.86	.95	.95	.54	.96 .91
Price-out	.98	.98	.98	.93	.61	.91	.88	.83	.98 .86
	.95	.97	.96	.90	.67	.95	.92	.85	.96 .89
Trade-off	.97	.98	.98	.96	.50	.76	.85	.73	.98 .75
	.95	.96	.95	.93	.56	.83	.90	.78	.95 .81

Table 4

Matching (r_m) and Pseudo-Matching Scores
Experiment II

Note. Correlations computed between the composites from the weighting scheme and the "true" model composites.

^aThe first row of each weighting scheme uses the composites.

^bThe second row of each weighting scheme uses the OLS-regression weights in forming the "true" model composites.

Equal weighting correlations: Model P, .93 (ridge) and .91 (OLS); Model N, .19 (ridge) and .30 (OLS).

Extreme weighting correlations: Model P, .93 (ridge and OLS); Model N, .89 (ridge and OLS).

subjects N-1 and N-2, all four post-training weighting schemes corresponded more highly to the criterion than did any of the statistical weights. For subject N-4, the statistical weights were better than the post-training weights. There were no differences for subject N-3, who showed the best matching and pseudo-matching of all the Model N subjects across all eight sets of assessed weights. For all four Model N subjects, the hybrid weights (ridge-ratio and OLS-ratio) composites demonstrated the largest divergence from the "true" criterion weight composites.

The pseudo-matching baseline correlatins for equal and extreme weighting, given at the bottom of Table 4, indicate that the Model P subjects' elicited weights were an improvement over either of the two heuristic weighting schemes. Although the equal weighting scheme for Model N is rather poor, the extreme weighting heuristic provides as high a pseudo-matching score as any of the Model N subjects, with the notable exception of subject N-3.

Discussion

Experiment II was designed to test the validity of eight procedures for assessing importance weights in a more complicated multi-attribute situation than that of the first experiment. The construct of "overall diamond value" was less predictable from the four attributes provided ($r_e = .82$ and $.83$ for Models P and N, respectively). Also, the set of alternatives, S, was constructed so as to present large intercorrelations among the four diamond attributes, (one large positive correlation for Model P, and two large negative correlations for Model N). The study is the first attempt to test the criterion validity of subjects' ridge weights, hybrid weights (suggested by Newman, Note 2), and indifference weights.

For Model P, all four subjects learned the diamond model well (high r_a), and provided weights consistent with holistic evaluations (high r_s), with one another (high convergence), and with the "true" MAU model taught (high r_m). Virtually no differences were observed among the eight sets of subjects' weights, in terms of the composites derived from them. Thus, the results of Model P subjects indicated that the criterion validity evidenced in Experiment I also holds in a context in which the weights are less explicit (lower R_e and non-zero attribute correlations) and over a broad range of weight assessment approaches. The success of the novel ridge and hybrid techniques is especially important.

The results for Model N are not consistent with those for Model P and Experiment I. One of the problems was that two of the four subjects did not learn the MAU appraisal model very well, as was evidenced by the low r_a scores. For the two subjects who did learn the model, weights were obtained from one subject which were highly valid, but the non-statistical weights obtained from the other subject yielded composites highly discrepant from those of the "true" diamond model. Although the weights obtained from the other two subjects were uniformly poor, there was a substantial superiority evidenced for the post-training assessments over the statistical and hybrid approaches.

The most surprising result is the extreme difference in Model P and Model N subjects' performance. The only difference between the two models is reflected in the sample intercorrelation matrix of the four attributes. Thus, subjects' ability to learn a MAU model (i.e., the relationship between attributes and an overall criterion construct of value) is greatly dependent upon the environmental

relationships among the salient attributes. Since only two subjects obtained satisfactory achievement scores, the results comparing assessment techniques for Model N are inconclusive.

Conclusions and General Discussion

Two experiments were conducted to assess the validity of several weight assessment techniques. In the first, a four-attribute MAU model with zero environmental correlations among attributes was taught to nine subjects. The regression-, rank-, and ratio-weight estimates all resulted in composites which closely matched those of the true model; most subjects' weighting schemes were a great improvement over either equal or extreme weighting. For three of the nine subjects, the rank and ratio assessments produced lower matching than did the regression-weight estimates.

In the second experiment, a total of eight subjects were taught one of two four-attribute MAU models, each involving substantial attribute intercorrelations. Both of these models were less explicit (more error variance) than the one taught in Experiment I. A total of eight methods were employed in assessing subjects' importance weights: OLS and ridge regression on holistic choices, OLS and ridge regression using ratio-weight estimates as validity coefficient estimates, direct subjective ranking and ratio estimation, and the two indifference techniques of pricing-out and trading-off to the most important dimension. For the model involving one large positive correlation between two of the attributes, all eight weight assessment methodologies produced equally good composites; all composites derived from subjects' weights corresponded to the "true" model composites better than simple heuristic rules such as equal weighting and extreme weighting. For the model involving two rather large negative intercorrelations among attributes, the results are inconclusive. Although the statistical weights were

superior for one subject who seemed to have learned the model well, the direct assessment and indifference weights were superior for two of the subjects who did not learn the model so well. Only one subject produced valid weights across all eight assessment techniques.

The present research and findings are interesting from both an applied and theoretical perspective. For the applied decision analyst (or judgement analyst), the work by Schmitt (1978) and that presently reported contribute strong evidence to the assertion that the additive MAU model is a valid prescriptive tool. The evidence that people can indeed provide direct subjective estimates of importance weighting is important. In most interesting decision problems, such as choosing a school desegregation plan or siting a nuclear power plant, a large alternative set is not readily known a priori. In such applied situations, the feasibility of most indirect holistic approaches to deriving importance weights is in doubt. Even if a reasonably large set of alternatives could be generated, in most cases the number of dimensions involved makes the task of holistic evaluation of alternatives extremely difficult, if not impossible.

The applied decision problem of Edwards (Note 3) is a good example. The decision-makers -- the board members of the Los Angeles Unified School District -- were faced with a MAU problem of seven alternatives. Each alternative was a detailed (or not so detailed) plan for desegregating the Los Angeles school system. In the final decision tree developed by Edwards, these plans were defined on 144 dimensions of importance. Any approach to defining the importance weights that depended upon holistic assessments of these few

alternative plans, defined on so many dimensions, would have been hopelessly inadequate. Ratio weighting, the assessment technique actually applied, was much more reasonable. The board members found the task of assigning ratio weights not only possible, but somewhat therapeutic. That is, Edwards' ratio-weight procedure forced them to think hard about their values and how they related to the overall utility of various desegregation plans.

Most applied decision analysts would like to think of themselves as more than therapists, however. The overwhelming belief among most decision analysts is that their methods elicit parameter estimates of preference models that result in a normative choice structure. That is, decision analysts believe that their clients should behave in the manner suggested as optimal by the elicited choice structure. Although the stimuli used in the present study (diamonds defined on four dimensions) and in Schmitt's (1978) study (graduate applicants defined on four attributes) are simplistic, and the acquisition of information about attribute importance is contrived (feedback learning), the results suggest that attribute importance is a valid psychological construct. That people can make accurate estimates of importance weights in the laboratory setting is certainly a necessary condition for their being able to do so in the more complex and emotional settings usually faced by a decision analyst and his/her clients.

From a theoretical perspective, the present study and that reported by Schmitt (1978) are highly relevant in the current debate over the extent to which people are aware of their own cognitive processes. In a recent article on the topic of verbal reports of mental process, Nisbett and Wilson (1977) summarize the Slovic and

and Lichtenstein conclusions on subjective weighting as a "fair assessment of this literature." In their review, however, Nisbett and Wilson used the term "impressive" in describing the "evidence of at least some correspondence between subjective and objective weights (p. 254)." Of course, Nisbett and Wilson's perception of the subjective weighting literature (based on the conclusions of Slovic and Lichtenstein, 1971) is out of date. Given the present data and the review by John and Edwards (Note 1), there is little reasonable justification for the claim that subjects cannot directly report beliefs about attribute importance.

From the perspective of Nisbett and Wilson, however, even the mostly negative conclusions of Slovic and Lichtenstein (1971) had to be reconciled with an overwhelming literature that people are totally incapable of introspection about cognitive processes. (For a rebuttal to the Nisbett and Wilson conclusions concerning self-insight and awareness in general, see Smith and Miller, 1978.)

Nisbett and Wilson (1977) assert the following:

It seems likely, in fact, that clinicians and stock-brokers could assign accurate weights prior to making the series of judgments in these experiments simply by calling on the stored rules about what such judgments should reflect. If so, one would scarcely want to say they were engaging in prospective introspection, but merely that they remember well the formal rules of diagnosis or financial counseling they were taught (p. 254).

The results from the current study, and Schmitt's (1978) study challenge Nisbett and Wilson's speculations. With only minor exceptions, subjects were able to provide importance weights predictive of their own holistic evaluations in an experimental setting for which there were no stored "rules" for determining judgments. The diamond appraisal policies in the present study were learned indirectly, without the intervention of verbal descriptions or formal linguistic rules. Subjects demonstrated an awareness of both their own rules for making diamond appraisals, and the criterion diamond model used to generate the outcome feedback.

The present study suggests that an important future variable in research on importance weighting is the intercorrelation matrix of attributes. Although the "true" criterion model is more difficult to determine when attributes are intercorrelated, the application of biased regression techniques makes the task a manageable one. The results of the present study were moderately encouraging for the novel hybrid weighting approach suggested by Newman (Note 2); further research is needed, however.

A possibly important intervening variable in the assessment of importance weights is the amount of exposure subjects have to the "true" MAU model. Also, the explicitness of the MAU model is another potential intervening variable. If the overall utility of the stimuli are not predicted well by the attributes considered (high error variance), subjective weights may not be so accurate. The amount of experience (number of learning trials) of the decision-maker, and the strength of the relationship between the MAU model attributes

and the construct of overall utility (\bar{R}_e), are concrete variables, often highly descriptive of specific applied settings. The first variable relates to the notion of decision-maker expertise, while the second is a function of the defining characteristics of the decision problem. Future research on importance weighting should systematically explore the effects of the number of trials of feedback learning and \bar{R}_e on subjective estimates of attribute importance. The problem of group assessment of importance weights is yet an additional topic for future research that has heretofore received little or no attention.

REFERENCE NOTES

1. John, R. S., & Edwards, W. Importance weight assessment for additive, riskless preferences functions: A review (SSRI Tech. Rep. 78-3). Los Angeles: University of Southern California, Social Science Research Institute, December, 1978.
2. Newman, J.R. Reflections on prediction models and decision analysis models. Unpublished manuscript, 1977. (Available from Social Science Research Institute, University of Southern California, Los Angeles.)
3. Edwards, W. Multiattribute utility measurement in a highly political context: Evaluating desegregation plans in Los Angeles. Unpublished manuscript, 1978. (Available from Social Science Research Institute, University of Southern California, Los Angeles.)

REFERENCES

- Keeney, R. L., & Raiffa, H. Decisions with multiple objectives.
New York: John Wiley, 1976.
- Nisbett, R. E., & Wilson, T. D. Telling more than we can know:
Verbal reports on mental processes. Psychological Review, 1977,
84, 231-259.
- Schmitt, N. Comparison of subjective and objective weighting
strategies in changing task situations. Organizational Behavior
and Human Performance, 1978, 21, 171-188.
- Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression
approaches to the study of information processing in judgments.
Organizational Behavior and Human Performance, 1971, 6, 649-744.
- Smith, E. R., & Miller, F. D. Limits on perception of cognitive
processes: A reply to Nisbett and Wilson. Psychological
Review, 1978, 85, 355-362.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director Advanced Research Projects Agency Attention: Program Management Office 1400 Wilson Boulevard Arlington, Virginia 22209	2 copies
Office of Naval Research Attention: Code 455 800 North Quincy Street Arlington, Virginia 22217	3 copies
Defense Documentation Center Attention: DDC-TC Cameron Station Alexandria, Virginia 22314	12 copies
DCASMA Baltimore Office Attention: Mr. K. Gerasim 300 East Joppa Road Towson, Maryland 21204	1 copy
Director Naval Research Laboratory Attention: Code 2627 Washington, D.C. 20375	6 copies

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, DC 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense
Research and Engineering (Test and
Evaluation)
The Pentagon, Room 3C125
Washington, DC 20301

Director, Defense Advanced Research
Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, ARPA Regional Office (Europe)
Headquarters, U.S. European Command
APO New York 09128

Director, ARPA Regional Office (Pacific)
Staff CINCPAC, Box 13
Camp E. M. Smith, Hawaii 96861

Dr. Don Hirta
Defense Systems Management School
Building 202
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum
Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, DC 20319

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, DC 20374

Vice Director for Production
Management Office (Special Actions)
Defense Intelligence Agency
Room 1E863, The Pentagon
Washington, DC 20301

Command and Control Technical Center
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Operations
(OP-951)
Washington, DC 20450

Office of Naval Research
Assistant Chief for Technology (Code 200)
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
493 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director, ONR Branch Office
Attention: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, DC 20380

Headquarters, Naval Material Command
(Code C331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Superintendent
Naval Postgraduate School
Attention: R. J. Roland, (Code 52R1)
C3 Curriculum
Monterey, CA 93940

Naval Personnel Research and Development
Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Naval Training Equipment Center
Human Factors Department (Code N115)
Orlando, FL 32813

Naval Training Equipment Center
Training Analysis and Evaluation Group
(Code N-00T)
Attention: Dr. Alfred F. Snodde
Orlando, FL 32813

Director, Center for Advanced Research
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code 540)
Attention: Dr. John Shore
Washington, DC 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

Department of the Army

Deputy Under Secretary of the Army
(Operations Research)
The Pentagon, Room 2E611
Washington, DC 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, DC 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and Systems
Research Laboratory
U.S. Army Institute for the Behavioral
and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director, U.S. Army Concepts
Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center
Attention: DRXMC-LS-SCAD (ORSA)
Ft. Lee, VA 23801

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Marine Corps Representative
U.S. Army War College
Carlisle Barracks, PA 17013

Chief, Studies and Analysis Office
Headquarters, Army Training and Doctrine
Command
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office
(Durham)
Box CM, Duke Station
Durham, NC 27706

Department of the Air Force

Assistant for Requirements Development
and Acquisition Programs
Office of the Deputy Chief of Staff for
Research and Development
The Pentagon, Room 4C331
Washington, DC 20330

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Deputy Chief of Staff, Plans, and
Operations
Directorate of Concepts (AR/XOCCC)
Attention: Major R. Linhard
The Pentagon, Room 4D 1047
Washington, DC 20330

Director, Advanced Systems Division
(AFHRL/AS)
Attention: Dr. Gordon Eckstrand
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffis AFB
Rome, NY 13440

IRD, Rome Air Development Center
Attention: Mr. Frederic A. Dion
Griffis AFB
Rome, NY 13440

HQS Tactical Air Command
Attention: LTCOL David Dianich
Langley AFB, VA 23665

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, DC 20505

Director, Center for the Study of
Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, DC 20505

Mr. Richard Heuer
Methods & Forecasting Division
Office of Regional and Political Analysis
Central Intelligence Agency
Washington, DC 20505

Office of Life Sciences
Headquarters, National Aeronautics and
Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, DC 20546

Other Institutions

Department of Psychology
The Johns Hopkins University
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Varial Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A. D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Brunel University
Attention: Dr. Lawrence D. Phillips
Uxbridge, Middlesex UB8 3PH
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Markhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Department of Electrical and Computer
Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Department of Government and Politics
University of Maryland
Attention: Dr. Davis B. Bobrow
College Park, MD 20747

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. Andrew P. Sage
School of Engineering and Applied
Science
University of Virginia
Charlottesville, VA 22901

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

Professor Raymond Tanter
Political Science Department
The University of Michigan
Ann Arbor, MI 48109

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Department of Psychology
University of Oklahoma
Attention: Dr. Charles Gettys
455 West Lindsey
Dale Hall Tower
Norman, OK 73069

Institute of Behavioral Science #3
University of Colorado
Attention: Dr. Kenneth Hammond
Room 201
Boulder, Colorado 80309

Decisions and Designs, Incorporated
Suite 600, 8400 Westpark Drive
P.O. Box 907
McLean, Virginia 22101

THIS PAGE IS BEST QUALITY PRINTED
FROM COPY FORWARDED TO DDC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 001922-6-T	2. GOVT ACCESSION NO. (Weights)	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Subjective Versus Statistical Importance A Criterion Validation.	5. TYPE OF REPORT & PERIOD COVERED Technical 10/77-12/78	6. PERFORMING ORG. REPORT NUMBER 78-7
7. AUTHOR(s) (10) Richard S./John Ward/Edwards	8. CONTRACT OR GRANT NUMBER(s) Prime Contract N00014-76-C-0074	9. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of So. Calif., L.A. CA90007	10. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22200	11. REPORT DATE (11) Dec 1978
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Decisions & Designs, Inc. 8400 Westpark Drive, Suite 600 McLean, VA 22101	15. SECURITY CLASS. (of this report) Unclassified	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited (12) 52p.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) (9) Research rept. Oct 77-Dec 78		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Multi-attribute utility, subjective weights, statistical weights, ridge regression, multiple-cue probability learning, outcome feedback, assessment procedures, criterion validity, environmental correlations		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The present paper proposes a research paradigm for comparing weight estimates to empirically derived "true" weights, thus obtaining a measure of the criterion validity of different weight estimation techniques. Subjects are first taught a multi-attribute utility (MAU) model via multiple-cue probability learning (MCPL) and outcome feedback. Then, various assessments of the importance weight parameters for the model attributes are obtained. Composites formed from these weights are subsequently compared to composites formed from optimal statistical weights derived from outcome feedback. → over		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

390.664 Ewu

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

↓ Data are reported from 17 subjects who were taught one of three "diamond worth" MAU models in 100 feedback trials. The models all involved four attributes (cut, color, clarity, and carat weight), and varied in the "environmental correlations" among the dimensions (either (1) all uncorrelated, (2) one large positive correlation, or (3) two large negative correlations). In addition to the usual MCPL indices of consistency, achievement, and matching, pseudo-matching correlations were computed for weights elicited via the direct subjective procedures of ranking and ratio estimation, the indifference procedures of pricing-out and trading-off to the most important dimension, and regression weights derived from subjective estimates of the validity coefficients. Overall, the composites formed from the subjects' elicited weights closely corresponded to the "true" weight composites. In addition, a high degree of correspondence was demonstrated among all of the assessed weighting schemes. Individual differences are also reported.

↖ The results of the present study are discussed from both an applied and theoretical perspective. To the decision analyst in the field, the present results give support to the belief that the parameter estimates obtained from clients define a "true" normative preference function. Theoretically, the findings of this study are strong evidence that people are aware of their own cognitive processes. ↗

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)